The history of the Polynesians inferred from high resolution HLA data.

Master thesis in human evolutionary genetics

Maja Krzewińska

Spring 2007

Centre for Evolutionary and Ecological Synthesis

University of Oslo, Norway

Table of contents

Summary

Numerous scientific investigations have been undertaken to answer the question of Polynesian origins. Most archaeological and linguistic evidence and previous studies in population genetics suggested that Remote Oceania was settled from the west. The human colonization of the Pacific began with the settlement of Australia and Papua New Guinea (PNG) at least 40,000 years ago, and was followed by the settlement of Remote Oceania, starting about 3,500 years ago. Most scholars agree that present-day Polynesians originated in Southeast Asia, and first entered the region as Neolithic agriculturalists. But the questions of the character of colonization and the extent of the Melanesian contribution to the Polynesian gene pool, as well possible South American contacts, still prevail. Certain plants, such as the sweet potato and bottle gourd, indicate some contacts with South America. Moreover, the Norwegian explorer Thor Heyerdahl showed that a voyage from South America was physically possible, even though no direct evidence of Amerindian ancestry has been found. Interestingly, recent research on the HLA system showed a significant South American genetic component in Easter Island.

In order to address the issue of possible ancient contacts between Polynesia and South America, and to investigate the genetic affinities of the Eastern Polynesians, 52 human genomic DNA samples from three locations in Eastern Polynesia (Southern Marquesas, Mangareva and Western Samoa) were tested for HLA polymorphisms. The results indicate that the Eastern Polynesians are principally derived from Southeast Asia. Some recent European admixture was detected, most likely due to post-European contacts. Several polymorphisms point to a possible North Asian connection (DRB1*0405 and DRB1*1201) and other, rare elements (e.g. A*0212, B*3905 and DRB1*1402) are clear evidence of South American admixture. No significant Melanesian ancestry was found. Phylogenetic analysis showed that the Polynesians studied did not differ significantly from Southeast Asian populations of mixed origins (such as Malay, Filipino, East Timorese and Moluccans) but have very little in common with the present-day inhabitants of Papua New Guinea. The results, however, could be the effect of sampling variance or genetic drift in Polynesia. In general, the HLA polymorphisms observed in this study give further support to the 'Express train' model of the colonization of the Pacific.

1. Introduction

1.1. Background

Over decades, the origin of the inhabitants of Remote Oceania has been the subject of much scientific scrutiny. The human colonization of the western Pacific (Australia, Papua New Guinea (PNG) and the Solomon Islands) has been dated to at least 40,000 years before present (Bellwood, 1978; Wickler and Spriggs, 1988). The descendants of those early settlers still inhabit present-day PNG and island Melanesia. They speak Papuan languages which have been shown to be of greater antiquity in the region than Austronesian languages. The latter language group is said to have been introduced into Oceania by Neolithic farmers. The appearance of horticulture in the region has been associated with the emergence of the Lapita culture characterized by notched pottery. The last wave of Polynesian settlement reached New Zealand about 800 years ago (Anderson, 1991).

Numerous scientific studies in the field of modern genetics have confirmed the Asian ancestry and significant homogeneity of present-day Polynesians. Albeit the Asian origin of the Polynesians is undisputed, there is evidence of possible South American contacts, such as plants of American origin (sweet potato, bottle gourd)(Borrell, 2007; Green, 2000). Another puzzling issue is the extent of contact between the proto-Polynesians and the earlier "Melanesian" settlers of Near Oceania.

Different theories have been brought forward to explain the colonization of the Pacific. Thus, the distribution of Austronesian languages and mtDNA analysis support an 'Express train to Polynesia' model of colonization (Diamond, 1988), according to which the expansion was recent and rapid. Conversely, globin gene polymorphisms (and some elements of Lapita culture) suggest that Polynesia was settled by people of South East Asian origin who established long-lasting and intense interactions in the region of Island Melanesia. This is so called the 'Entangled Bank' model of colonization (Terrell, 1988).

Modern genetics enables us to reveal more of the human past than it was ever possible. One of the systems used in the field is the Major Histocompatibility Complex (MHC), which plays a vital role in the immune response. The complex, known in humans as the Human Leukocyte Antigen system (HLA), is useful due to its high polymorphism, bi-parental inheritance, and the fact that it is less subject to lineage extinctions than the uniparental mtDNA and Y chromosome lineages. HLA analysis, due to the high variability of the complex and rapid technical development of high throughput and high resolution typing

methods, is capable of distinguishing between highly diverse alleles, and has proven to be a powerful tool in human population genetics (Bugawan et al., 1999).

In this thesis, I describe the analysis of HLA alleles in three Polynesian populations (Southern Marquesas, Mangareva and Samoa), to establish the extent of gene flow from other geographical regions. This study included 52 human DNA samples, with 20 samples from Southern Marquesas, 15 from Mangareva and 17 from Western Samoa (currently Samoa). The genetic data were obtained using high resolution HLA typing, and consisted of alleles from 5 loci (HLA-A, HLA-B, HLA-C, HLR-DRB1 and HLA-DQB1).

In the following section, I shall describe the structure of the HLA complex and give an account of its usefulness in population genetic studies.

1.2. The human leukocyte antigen complex

Human Major Histocompatibility Complex (MHC) in humans is called Human Leukocyte Antigen Complex (HLA). It is situated in a continuous stretch of DNA on the shorter arm of chromosome 6. The classical HLA complex covers around 3.6 Mb, with 224 described genes of which 128 are thought to be expressed. The extended HLA complex, containing parts of centromeric and telomeric regions, covers a total of 8 Mb. The complex is responsible for the development of immune responses, susceptibility to diseases and autoimmune responses. It is the most polymorphic region in the human genome.

HLA genes are organized into regions and encode three different classes of genes: class I HLA genes (HLA-A, HLA-B, HLA-C), class II HLA genes (HLA-DR, HLA-DQ, HLA-DP) and class III HLA genes (secreted proteins associated with immune responses e.g. complement (C') proteins and the tumour necrosis factors,TNF-α and TNF-β) (Kuby, 1997). The HLA complex is the most gene-dense region of the human genome.

The structure and function of both HLA class I and II molecules are closely related. Both are membrane-bound glycoproteins presenting antigen molecules on the cell surface. They form stable complexes with antigenic peptides on the cell surface for T-cells to recognize. Both HLA classes form three- dimensional structures.

1.2.1. Organization of genes encoding class I and II molecules

Each domain of the MHC molecule is encoded by separate exons. Class I genes are usually encoded by a 5' leader exon (encoding a signal peptide) followed by 5-6 exons encoding the α-chain of the molecule. For HLA class I typing purposes, exons two and three are usually sequenced.
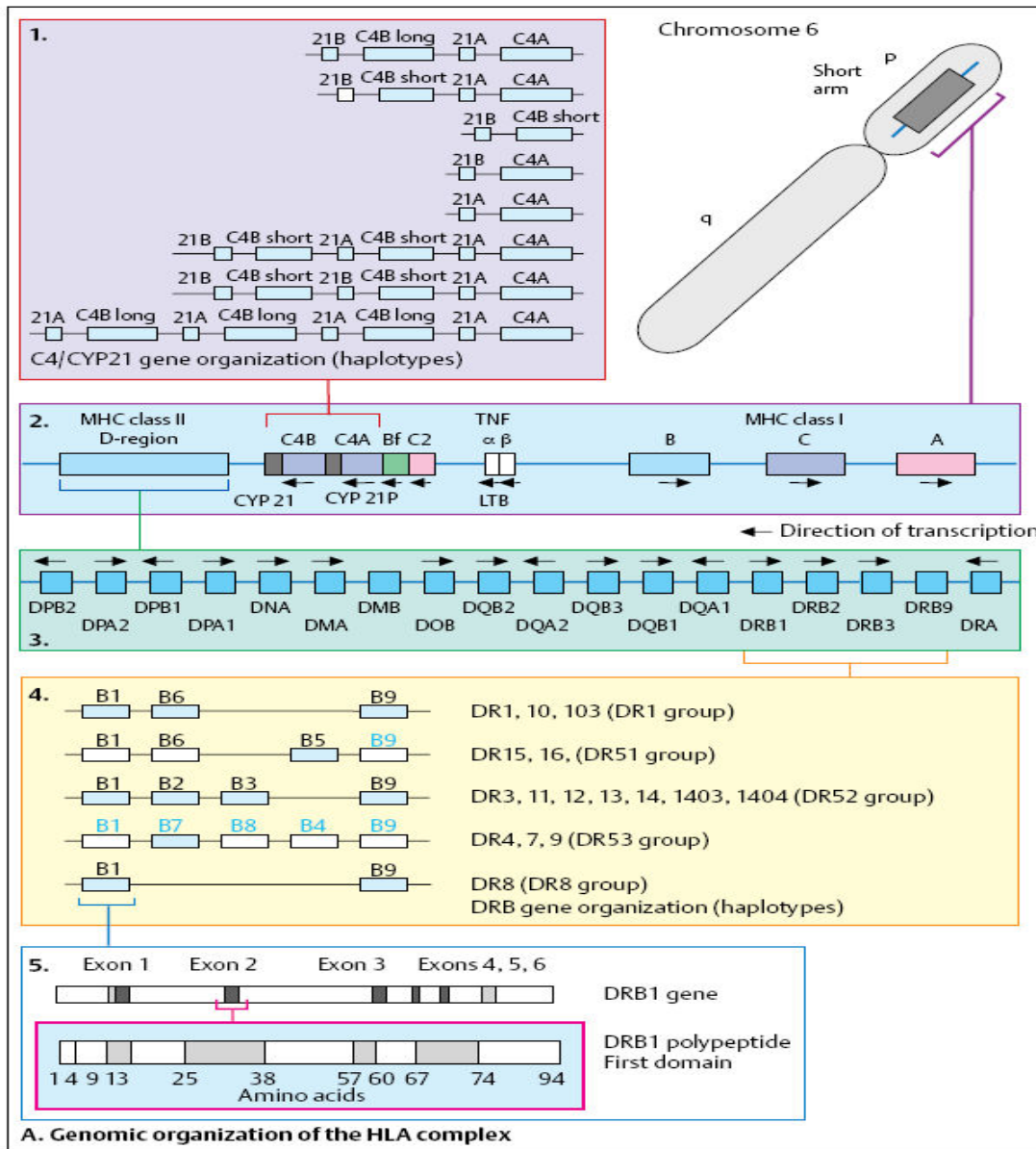
Figure 1 Genomic organisation of HLA complex and its localisation on the shorter arm of chromosome 6 (Brumester and Pezzutto, 2003). The expression products of complementing factors are called class III antigens (box 1). Genes encoding class III molecules are located between class I and class II MHC genes (box 2). Structurally related genes are located in the vicinity of DP, DR and DQ genes (box 3). Class II molecules are encoded by different gene loci which number and structure differs from one individual to next depending on haplotype. Here different combinations of β–chain encoding regions are shown to result in different DR antigens (box 4). Gene configuration: the gene is divided into exons encoding for numerous domains (box 5).

In the genes encoding Class II molecules, exons are organized into series that correspond with the domain structure of the two chains (α- and β-chain). The genes start with a 5' leader exon that is followed by 4-5 exons encoding the α- and β-chains of the molecule (Kuby, 1997). For HLA class II typing purposes (HLA-DRB1 and HLA-DQB1), exon two of the respective gene is usually sequenced.

1.3. Nomenclature of HLA complex

The nomenclature of HLA is directly related to the typing method used. Thus, serological nomenclature consists of letter coding, describing the class type a given molecule belongs to (gene locus), and numeric characterization of the antibody recognizing the HLA molecule (e.g. DR4, A10). Traditionally, the letters were assigned to the different loci according to the order in which the loci were discovered (A, B, C and D) and the number represented the individual allele within the group. But with time, as the specificity could be more precisely defined (e.g. A10 could spilt into A25(10) and A26(10)), the nomenclature became too confusing and had to be replaced with new nomenclature (Browning and McMichael, 1996; Brumester and Pezzutto, 2003).

Introduction of classification based on allelic definition resolved the problem of HLA nomenclature. Alleles defined by DNA typing (genomically defined alleles) are named after locus name and followed by an asterisk and a number indicating the particular allele (e.g. DRB1*0401). The number used usually consists of four digits where the first two digits define specificity and the following digits are the allele assigned within this specificity (e.g. A*0205 defines allele 5 of HLA-A2) (Browning and McMichael, 1996). The name can consist of up to 8 digits (e.g. HLA-DRB*13010102). In such cases, a six-digit name describes alleles differing by synonymous mutations and eight-digit names describe alleles containing mutations outside the coding region.

1.4. Polymorphisms

MHC loci are highly polymorphic (multiple alleles on each genetic locus).The polymorphisms are clustered within sequences encoding for a peptide-binding groove and tend to be non-synonymous. In other words, they tend to alter the sequence or conformation of the peptide (Browning and McMichael, 1996). MHC loci are also closely linked and therefore each individual inherits two sets of alleles – one from each parent. Such a set of alleles is called a haplotype (Kuby, 1997). The HLA alleles are co-dominant (both alleles are functional and are translated into proteins). The alleles on each genetic locus can differ by up to 20 amino-acid residues (sequence divergence of 5-10%). It has been shown that polymorphic differences are not unique sequence motifs. *'Instead, most alleles gain their uniqueness through a mosaic of conserved polymorphic differences, each defined at a particular hypervariable region, resulting principally from gene conversion, recombination and exon shuffling events'* (Browning and McMichael, 1996).

Existence and maintenance of high HLA gene polymorphism in populations is influenced by several agents such as population size, non-random mating, mutation, selection (e.g. as a response to pathogenic pressure (Browning and McMichael, 1996)), recombination or gene conversion (e.g. from pseudogenes) and random drift.

1.5. Inheritance

Independent inheritance, as described by Mendel, is possible only with genes located on different chromosomes. The MHC alleles are co-dominantly expressed meaning that both haplotypes (maternal and paternal) are expressed equally in the descendant. In an outbred population there is a one in four chance that siblings will inherit the same parental haplotypes (if different) and become histocompatible (Kuby, 1997). In other words, since these genes are tightly linked they tend to be inherited as a haplotype. This haplotype-specific inheritance enables population structure analysis and can lead to inference of genealogies from known haplotypic structure.

1.6. Linkage disequilibrium and recombination

One of the most important factors influencing the inheritance of HLA genes is the phenomenon of linkage disequilibrium. If two linked loci are occupied more often by alleles frequently occurring together (associated) on a haplotype than expected by chance, the situation is often referred to as linkage disequilibrium (Hedrick and Kumar, 2001). The phenomenon of linkage disequilibrium is also known as gametic association (Browning and McMichael, 1996) or co-segregation of loci. A number of factors such as gene flow, distance between the loci, migration, directional selection, population growth or mutation can be responsible for linkage disequilibrium.

Co-segregation heavily depends on the recombination fraction (θ) that uses recombination frequencies as its measure. The relation between the recombination fraction and linkage disequilibrium is rather complicated and recombination hotspots have been observed between loci in strong linkage disequilibrium. Recombination hot spots have been observed in classical HLA complex, DQ-DP region and B-TNF section. However, no crossover events have been described in the intervals between B-C and DRB1-DQA1 which may reflect selection for particular alleles and allele combinations (Browning and McMichael, 1996).

The measure of linkage disequilibrium is the degree of linkage disequilibrium (D). It is the difference between observed and expected haplotype frequency based on individual allele frequencies (Hedrick and Kumar, 2001).

Both recombination rate and linkage disequilibrium differ throughout genome and the degree of linkage disequilibrium is particularly high in the HLA complex causing the alleles to often occur in particular combinations.

1.7. Population genetics

As mentioned above, each individual inherits two haplotypes – one from each of the parents. The more outbred the population, the more heterozygous individuals it will have (each specimen will generally be heterozygous at each locus). In contrast, the more inbred the population, the more homozygotes are observed (the less different haplotypes are present in the gene pool of the population)(Kuby, 1997).

High allelic diversity of HLA genes is useful in population studies. Population data can be used to help understand human evolution and migration patterns as well as the selective forces operating on HLA loci, e.g. restricted diversity on some of the loci could be consistent with population bottlenecks in founding populations (Browning and McMichael, 1996).

Some allelic combinations between class I and class II molecules or within the classes have been proven to represent haplotypes specific for particular ethnic groups. Such population-specific, unique alleles can be used in search for ancestral alleles and therefore are very informative in population genetics. Population studies can reveal new alleles, and the operation of selective forces on the HLA polymorphism, such as symmetric balancing selection or positive directional selection for a specific allele (Bugawan et al., 1999). Symmetrical balancing selection is postulated to operate on DR and DQ allele frequencies (Browning and McMichael, 1996)

2. Genes and Polynesian history

2.1. History and genes

Genetically, Polynesians are most similar to Asian and other Pacific populations and quite different from Native Americans, which appears to reject Heyerdahl's hypothesis of a New World origin of Polynesians (Relethford, 2003). But if Polynesians actually came from Asia, where should we search for their origins? As described in the following section, linguistic analysis points to Taiwan and southern China as the place of origin. Archaeology based on Lapita findings, points to the southern coast of China. Some elements of the culture are even said to point to Melanesia as the birth place of Neolithic voyagers. Out of the three locations, Taiwan seems to have greatest scientific support as the starting point. However, Su and colleagues examined Y-chromosome haplotypes in Taiwan, Southeast Asia, and Polynesia (Su et al., 2000) and found that the Polynesian Y chromosomes were a subset of Southeast Asian Y chromosomes, and differed significantly from the Taiwanese Y chromosomes. In this case, both Taiwanese and Polynesian populations should be classified as descendants of a Southeast Asian population, but Taiwan should not be viewed as the place of origin of Polynesians. This demonstrates that genetic inferences about the human past are complicated and should use as much available information (archeo-linguistic evidence) as possible. Even then, the analysis might be incomplete due to the fact that the genetic composition of modern-day Polynesian populations must have changed over time. Therefore, it is difficult to analyze ancestral human populations on the basis of modern human populations. Many human lineages might have disappeared over time and therefore modern people should not be treated as accurate reflection of their ancestors.

2.2. The Pacific Islands and Polynesia

The islands of the Pacific are usually referred to as 'Near Oceania' and 'Remote Oceania' and are traditionally divided into Polynesia, Melanesia and Micronesia. Near Oceania includes such islands as New Guinea, New Ireland and New Britain (Bismarck Archipelago), all of which were first settled between 40,000-11,000 years ago by hunters-gatherers that came from the west (Bellwood, 1978; Diamond, 1988).

Further to the east, the first evidence of occupation appears around 3,500 years ago. 'Inter-visibility' of land masses is said to have played a very important role in the colonization of Remote Oceania. The colonization of the Pacific is associated with the appearance of the

Lapita culture in the region and a change in the ecology of the islands (habitat modification, introduction of domesticated plants and animals, and hunting animals to extinction, e.g. the giant moa of New Zealand).

Figure 2 Map of the Pacific with marked positions of the islands in the study (http://faculty.washington.edu/ plape/ pacificarchwin06/pacificarchschedule.htm from Keegan, William F. and Jared Diamond (1987) Colonization of islands by humans: a biogeographical perspective. *Advances in Archaeological Method and Theory* 10: 49-92).

2.2.1. Mangareva

Mangareva is located in French Polynesia and is the central and most important island of the Gambier Islands. It is surrounded by other smaller islands: Taravai in the southwest and Aukena and Akamaru in the southeast and other smaller islands also located in the north. It covers approximately 18 km^2 (about 56% of the land area of the whole Gambier group). Rikitea, which is the largest village on the island, is also the chief town of the Gambier Islands. Mangareva was once heavily forested and supported a large population that traded with other islands via ocean-going canoes. However, excessive logging by the islanders during the tenth to fifteenth centuries resulted in deforestation of the island, with disastrous results for environment and economy. Mangareva has been first settled from the north by settlers from Southern Marquesas around 1,200 A.D. (Jennings, 1979), and its population was estimated to 860 inhabitants during the 1996 census.

2.2.2. Marquesas

The Marquesas also lie in French Polynesia. The archipelago consists of two groups of islands: Northern Marquesas (Eiao, Hatutu, Matu Iti, MatuʻOa, Matu One, Nuku Hiva, Ua Huka, Ua Pu) and Southern Marquesas (Fatu Hiva, Fatu Huku, HivaʻOa, Moho Tani, Motu Nao, Tahuata, Terihi).

The group is formed by peaks of an elevated submarine volcano. The islands were discovered by the Spanish navigator Alvaro Mendana in July 1595. Almost 200 years later, in 1774, Cook rediscovered the southern islands, and in 1791 the northern group was sighted by Joseph Ingraham. The first settlers came from Samoa to the northern island group around A.D. 300 and spread to the southern group some time after A. D. 600 (Jennings, 1979). Pre-European history of the islands can be divided into four distinct cultural phases, but generally archaeological material shows that inhabitants fishermen and farmers living in the valleys (Bellwood, 1987). It is not clear whether they were also pottery makers since the analysis of the very few potsherds discovered suggested that pottery might have been imported from Fiji or Tonga (Jennings, 1979). The role of Marquesas Islands in early Polynesian prehistory is fairly well supported by archaeological evidence and it has been shown that the island group was most probably a dispersal centre for the rest of East Polynesia (New Zealand, Society Islands, Mangareva, Henderson and Pitcairn). The pre-European population of Marquesas was estimated at 80,000 inhabitants, but the number was drastically reduced until 1926 when it reached a critical 2,000. The 1970 census population was estimated to have increased to 5,400 (Jennings, 1979).

The island group had been provisioning place for whalers since the 1790's, and the export of sandalwood since 1815 caused much interest in the region. The first missionaries came in 19th century. The island group was annexed by France in 1842 (Campbell, 1989)

2.2.3. Western Samoa

Western Samoa is officially known as the Independent State of Samoa or Samoa, and was called Western Samoa between 1914 and 1997. Before that time (between 1900 and 1919) it was often referred to as German Samoa and Navigator Islands (before the 20th century). The population of Samoa was 176,710 inhabitants in 2001. Samoa is an island group of volcanic origin and its two largest islands (Savai'i and Upolu) lie in the western part of the archipelago. The group was first discovered by Roggeveen in 1722 and Bougainville (1768) but neither of them came ashore. The first European to visit the islands was La Pérouse (1787)

(Fischer, 2002). Samoa was probably settled together with Tonga around BC 1,500. Samoan prehistory since about AD 200 has been aceramic (Bellwood, 1987).

Contact with Europeans intensified after the 1830s, when English missionaries and traders began arriving in the region. Mission work in Samoa had begun in late 1830 by John Williams, of the London Missionary Society (Campbell, 1989). By the late nineteenth century Samoa was valued by the French, the British, the German and the American merchants as a refuelling station for coal-fired shipping (Fischer, 2002). As Germany began to show more interest in the Samoan Islands, the United States laid its claim to them. Britain also expressed its interest. By the end of 19th century, the islands were split into two: the eastern group became a territory of the United States and are today known as American Samoa; the western islands became known as German Samoa (until 1914). Between 1918 and 1962 Samoa was controlled by New Zealand (Campbell, 1989).

2.3. Where did the Polynesians come from?

2.3.1. Linguistic evidence

The languages spoken in the Pacific region are usually divided into two different groups: Austronesian and Papuan languages. The former group numbers about 1,200 languages which form ten subgroups: nine of these subgroups are only found in Taiwan (26 languages), and the tenth subgroup (Malayo-Polynesian) includes all the remaining 1,174 Austronesian languages, which are spoken outside Taiwan (from Madagascar to east Polynesia)(Diamond, 2000). The Polynesians speak over 30 closely related languages forming a minor branch (Oceanic subgroup) of the whole Austronesian language family.

The second language group (Papuan) consists of languages spoken in New Guinea, large parts of the Bismarcks and Solomons. These languages probably descend from the languages of the first settlers that came to the region same 40,000 years ago and are unrelated to Austronesian (Bellwood, 1987). Papuan languages represent a group of languages belonging to different language families and it is not known if they share a common proto-language or not. But they do show a high degree of structural similarity which enables us to distinguish this group from Austronesian language family (Dunn et al., 2005).

Early language studies often concentrated on finding correlations between language patterns and observed anthropological characteristics of the natives. Based on language analysis, it has been suggested that the first inhabitants of Near Oceania spoke non-Austronesian languages and that Papuan languages are descendants of those early languages.

14

Moreover, Early Austronesian speakers have been identified as the source of Polynesian genes. However, further studies of the Polynesians showed that in Melanesia, for instance, many Austronesian-speaking populations are of Australoid origin and that there is no one-to-one correlation between Mongoloid populations and Austronesian languages (Bellwood, 1987).

The reconstructed vocabulary of Proto- Malayo-Polynesian language represents agricultural societies that grew rice, made pots, lived in well-built timber houses and kept domesticated animals. It contains words such as taro, breadfruit, banana, yam, sago and coconut; and others, such as the terms for pottery, sailing canoes and several components of timber houses (Bellwood, 1991). Traditionally, both archaeologists and linguists point to Taiwan or southern coast of China as the places of origin from which Austronesian expansion started. From there, the expansion continued to the Philippines, Indonesia and the Pacific Ocean.

2.3.2. Farming and colonization

The Austronesian expansion began about 5,000 years ago from Taiwan into the Philippines and Indonesia. The first artefacts to appear in the western Pacific can be dated to between 3,500 and 6,000 years ago. These artefacts, which include pottery, polished stone adzes, knives, spear points etc., are of Chinese type. From that period, some evidence for rice cultivation and for inland forest clearance is also available (Bellwood, 1991).

Vocabulary reconstructions show that taro, yams, banana, breadfruit and rice were cultivated in Island South-East Asia as early as 2,500 BC, and that the people domesticated pigs, dogs and chickens. The material culture of those early settlers consisted of pottery, outrigger canoes with sails and polished stone tools. This suggests that the Polynesians entered the Pacific from South eastern Asia as Neolithic horticulturists and fishermen and that the process of colonization was greatly facilitated by canoe-building and sailing techniques (Bellwood, 1987).

The pattern of the spread of Austronesian speakers differed in remote Oceania and in other islands which were already inhabited. It was a complex process accompanied by gene exchange in places where the newcomers met already well established and dense populations with developed agriculture (New Guinea, Northern Melanesia)(Bellwood, 1991). In contrast, the colonization of the Pacific and Micronesia by Neolithic settlers was an expansion into empty, previously uninhabited areas, unaccompanied by gene flow.

2.3.3. Lapita culture

The Lapita culture is the type of culture associated with characteristic notched pottery, domesticated animals and navigational skills. On the basis of 'proto-Oceanic' language reconstruction and especially reconstruction of the vocabulary related to oceanic voyaging, it has been suggested that Lapita pottery makers also spoke languages that belonged to the Oceanic languages (Jobling et al., 2004). The Lapita culture is related to a similar one that existed in Taiwan and South China about 6,000 years ago but some of its elements have been derived from more local practices.

The 'Lapita period' lasted from between 3,600-2,500 years ago (Bellwood, 1987) and it has been generally assumed that the Lapita homeland and the proto-Oceanic homeland was one and the same. It the remote Pacific, Lapita can be identified as an element of colonization process. Most probably, no second wave of colonization followed those first settlers and we can thus assume that present-day Polynesians (and their languages) are direct descendants of the Lapita settlers (Irwin, 1992). After the initial colonization, Polynesians continued travelling between distant archipelagos.

2.4. About migrations, 'slow' and 'fast boats'

The prehistoric colonization of the Pacific was a **two-phase process**. The first people to arrive in the region were the Pleistocene hunter-gatherers. The second phase of colonization consisted of Neolithic farmers who expanded from China and Taiwan around 6,000 BC (Gray and Jordan, 2000; Oppenheimer and Richards, 2001) and further to the Pacific. In order to explain the second process, different models of colonization of the Pacific by Austronesian speakers have been suggested. Two main contrasting hypotheses are: the 'Express train' and the 'Entangled-bank model' (Gray and Jordan, 2000). The theories account for different degrees of the extent and nature of contact between the newcomers and the Melanesians already inhabiting parts of the region.

The **'Express train model'** (Diamond, 1988), also called 'Out of Taiwan' model, assumes that the second Neolithic expansion was very rapid, that it passed by Melanesia and that it took the newcomers as little as 2,100 years to reach the western Polynesia outliers. The Austronesian expansion was possible because of the outstanding navigational skills of the newcomers (a fact supported by the archaeological evidence of obsidian exchange and its transport over vast distances), and it was rooted in the development of agriculture (Bellwood, 1991).

The **'Entangled bank model'** (Terrell, 1988), on the other hand, represents a slightly different model of expansion, in which more emphasis is put on interactions between the Pacific peoples, Melanesians and the inhabitants of South East Asia, than on colonization itself. In this theory, the observed similarities between Polynesians and Melanesians are probably the result of significant interbreeding in northern Island Melanesia (Irwin, 1992). This model emphasizes the role of Near Oceania, as a proximate origin of the inhabitants of Remote Oceania.

Another model of colonization, proposed by Kayser *et al*. (Kayser et al., 2000) is the **'Slow boat model'**, which assumes that the ancestors of modern-day Polynesians originated in China/Taiwan but mixed extensively with Melanesias on their way to Remote Oceania. This resulted in appearance of Melanesian genes in Polynesian gene pool, as well as introduction of Polynesian genes to Island Melanesia.

The three models are not mutually exclusive. And even though the language tree fits 'the express train model' best, archaeological and genetic data suggest that population interaction has probably occurred even between very distant archipelagos (Gray and Jordan, 2000), yielding support to the other hypotheses.

It is worth mentioning that some scholars have somewhat different views of Polynesian origins. According to Oppenheimer and colleagues, Polynesian origins could be somewhere between New Guinea and the Wallace's line (Oppenheimer and Richards, 2001). Other models are intermediate and differ in the respective relative contributions of Southeast Asian and Near Oceanian genes to the settlers of Remote Oceania. Among those, Jobling *et al.* mentions the 'Slow Train' and 'Triple I' models.

2.5. Genetic evidence – mitochondrial DNA and Y chromosome – 2 different (his)stories?

2.5.1. Classical markers

The analysis of classical markers in Polynesia did not give unanimous results on the origins of Polynesian peoples. On the contrary, Cavalli-Sforza's genetic tree of 31 populations from Pacific Islands constructed using genetic distances shows significant deviation from the classical classification of Pacific islands into Polynesia, Melanesia and Micronesia. This could be the result of high levels of post-settlement gene flow between Near and Remote Oceania. However, the author observed some clustering among Eastern Polynesians (Society, Cook, New Zealand and Easter Island) and few similarities between Australia and New Guinea. The patterns observed have been interpreted as a result of post-settlement, region-specific migrations. In general, the analysis of classical markers provides little information about settlement of Remote Oceania but clearly excludes possible South American origin and proves it very unlikely (Cavalli-Sforza L. L., 1996).

2.5.2. Globin gene polymorphisms

Deletions of the α-globin genes resulting in α-thalassemia have proved to be informative markers in the Pacific. They tend to be geographically specific. Globin deletions appear to protect against malaria and are present at notable frequencies amongst Pacific populations living in areas affected by the disease (e.g. New Guinea to Vanuatu) that is absent from locations in Remote Oceania. Polynesians carry haplotypes that are common in other populations but not Near Oceania (sometimes called 'Mongoloid' αα haplotypes) and others that are common in Near Oceania but absent or very rare in other populations (e.g. the -α$^{3.7}$III deletion). This is thought to be evidence of a Melanesian contribution to Polynesian gene pool and the dual origin of inhabitants of Remote Oceania (O'Shaughnessy et al., 1990).

2.5.3. Mitochondrial DNA

Mitochondrial DNA (mtDNA) diversity in Remote Oceania is most often based on variation observed in the hypervariable region I. In addition, in a section of noncoding mtDNA (between the cytochrome oxydase II (COII) and the lysyl transfer RNA genes) there is a specific 9-bp sequence (CCCCCTCTA). This sequence in some people can be doubled (forming a specific 18-bp sequence) and in other people one of the two adjacent copies of a 9-bp sequence can be missing, forming a marker known as 9-bp deletion. This deletion is

abundant in present-day Asian populations (5-40%) (Hagelberg et al., 1994) and often used as a genetic marker for East Asian ancestry, since it has been found in almost all East Asian populations, Native Americans, Polynesians and also in Pygmies from Africa (Redd et al., 1995). The deletion frequency in Aboriginal Australians is less than 1% (Redd and Stoneking, 1999).

In Polynesians, the deletion is often accompanied by 3 specific nucleotide substitutions in mtDNA sequence (at positions 16217, 1647 and 16261), giving a characteristic pattern known as the **'Polynesian motif'** (Hagelberg and Clegg, 1993; Melton et al., 1995; Redd et al., 1995). According to Alan Redd and colleagues, the motif becomes more and more prominent in Polynesian populations, reaching a peak of frequency in Samoa. It has been shown by Hagelberg and colleagues that Samoans carry either the 'Polynesian motif' or a haplotype which is directly ancestral to the motif (Hagelberg et al., 1999). The motif is absent in highland New Guinea and among Australian Aborigines (Redd et al., 1995; Relethford, 2003), nor was it detected in America suggesting that there was probably lack of significant prehistoric contact between Polynesia and the Americas (Hagelberg et al., 1994).

Other mtDNA lineages found in Remote Oceania tend to be located either in Island Southeast Asia but not in highland New Guinea or the other way around (e.g. mtDNA lineages P and Q). The existence of such a clear-cut distinction between observed mtDNA lineages enables discussion about population origins and seems to favour Diamond's 'Express train model' as the most likely model of colonization (Hagelberg et al., 1999). Some mtDNA types shared by Polynesians and Native American populations are also found in Asian populations, but there is no evidence of Native American contribution to maternal lineages on the remote island of Rapa (Hurles et al., 2003).

2.5.4. Y chromosome

A few Y chromosome lineages have been described in Polynesia. Additional European admixture in last 250 years (mainly by sailors, whalers, traders, missionaries, etc.) might obscure the picture of the original settlement. But this gene flow is easy to identify, since there are different degrees of admixture (mostly from northwest Europe) on different islands. The remaining lineages must therefore be paternal Polynesian lineages. The predominant lineage in Remote Oceania is haplogroup C which can be traced back to Island Southeast Asia (Jobling et al., 2004). Cook Islanders appear to carry just 3 Y-chromosome haplotypes. The DYS390.3del/RPS4Y711T haplotype was found in 82% of Cook Islanders studied. The same haplotype is also present in other populations such as Indonesian and Melanesian, but

surprisingly it is completely absent from populations in East and Southeast Asia and Australia (Kayser, 2000). This **'Oceanic motif'** (characterized by deletion within microsatellite DYS390, unique allele at the minisatellite MSY1 and an additional SNP, M38) exhibits the greatest variation within Wallacea where it occurs together with other haplogroup C chromosomes. Therefore, Wallacea is a likely candidate for place of origin of this 'Oceanic motif'. According to those findings, the best model to explain the colonization of Polynesia is the "slow boat" model in which Austronesian peoples moving eastwards into Polynesia 'picked up' some Melanesian genes on the way.

Native American Y chromosomes detected in the island of Rapa were explained to be the result of 19th century Peruvian slave trade in Polynesia rather than prehistoric migration and settlement from South America (Hurles et al., 2003). Generally, Polynesian migration seems to show sex-dependant patterns. MtDNA points back to Asia and Taiwan as the point of origin of Polynesian populations, and the Y chromosome points to Melanesia.

3. Materials and methods

3.1 DNA samples

The samples used in this thesis were obtained from Prof. Erika Hagelberg and were a gift of Prof. J.B. Clegg, Weatherall Institute of Molecular Medicine, University of Oxford, UK. The DNAs were previously used in studies of globin polymorphisms (O'Shaughnessy et al., 1990). The sample set consisted of 52 human genomic DNA samples: 20 samples were from Southern Marquesas (French Polynesia), 14 from Mangareva (French Polynesia) and 17 from Western Samoa. The DNAs were extracted from blood samples using a phenol/chloroform extraction method. Most of the samples had not previously been subjected to high resolution HLA typing.

3.2 Sample processing and concentration

3.2.1 Whole DNA amplification

The DNA samples were first treated with a "GenomiPhi DNA Amplification Kit" (Amersham Biosciences) (Table 3) in to amplify the amount of genomic DNA, which was later used for HLA typing. The protocol was the standard protocol supplied by the manufacturer. Before the procedure, the DNA concentration of the samples was measured, and found to range from 0.75 ng/µl (sample MA057, which later failed to yield results) to a maximum of 264 ng/µl (sample SM021) (see Results chapter).

3.2.2 Amplification of HLA loci

The samples were typed for five HLA loci (HLA-A, HLA-B, HLA-Cw, HLA-DRB1 and HLA-DQB1) polymorphisms using high-resolution sequence based typing (SBT) (Sayer et al., 2004b). In brief, PCR amplification with locus-specific amplification primers (for the list of the primers see Appendix; Table 1) was performed. The samples were sequenced on an ABI3730 Sequence Analyzer. The results were interpreted using ASSIGN-SBT software (Conexio Genomics, Applecross, Australia)(Sayer et al., 2004a).

For each locus, different sets of primers, PCR reagents and PCR programmes were used. The primers and PCR reagents are listed in Table 2 and Table 3 respectively (Appendix). In brief, HLA-A, HLA-B and HLA-C were amplified in total volumes of 25µl (the template DNA being 5µl) with 0.25µl of each amplification primer designed for

respective loci. For each locus one master mix was prepared containing both forward and reverse primers. The programmes chosen for PCR amplification differed significantly but they all started with one denaturation cycle of $96^{o}C$ (5-6min), followed by 30 cycles of 96-$98^{o}C$ (10-30s), $60-70^{o}C$ (30s) and $72^{o}C$ (1-2 min). The last step was prolonged extension cycle of $72^{o}C$ (10 min). The PCR finished with $4\,^{o}C$ according to the protocols developed in house.

HLA-DRB1 PCR amplification was performed in total volume of 12.5µl per sample (with 2.5µl of template DNA) containing different amounts of all 8 amplification primers (Sayer et al., 2004b). The PCR programme of choice started with one cycle of $95^{o}C$ (15 min) and was followed by 30 cycles of $95^{o}C$ (20s), $62^{o}C$ (10s) and $72^{o}C$ (90s). The finishing extension cycle was longer and lasted 8 min ($72^{o}C$) followed by '$4^{o}C$ forever' option. Finally, for PCR amplification of HLA-DQB1 locus two different master mixes were prepared because of two primer sets containing two different reverse primers. The total volume of each reaction was 15µl with 3µl of template DNA. The PCR programme used was the same for both amplification mixes (regardless of the reverse primer used) and consisted of one cycle of $96^{o}C$ (15 min), followed by 35 cycles of $96^{o}C$ (20s), $66^{o}C$ (10s) and $72^{o}C$ (90s). The cycles were succeeded by another 5 rounds of $96^{o}C$ (20s), $59^{o}C$ (10s) and $72^{o}C$ (90s) and the programme finished with $4^{o}C$.

All PCR amplifications were performed according to techniques and protocols used at the Immunology Institute at Rikshospitalet-Radiumhospitalet Medical Center, and detailed protocols can be made available on request.

The product length was determined by gel electrophoresis on 1 % agarose for HLA-C, HLA-B on 1% agarose gel (GeneRuler[TM] 1 kb; Fermentas), 1.5% agarose gel for HLA-A and HLA-DRB1 (GeneRuler[TM] 1 kb and 50 bp respectively) and 2% agarose gel for HLA-DQB1 (GeneRuler[TM] 50bp). Electrophoresis was performed at 80 V for 40min in 1xTBE buffer.

3.2.3 Sequencing of amplified products

All the samples that showed positive PCR amplification underwent Exo/SAP purification (using Exonuclease I and Shrimp Alkaline Phosphatase). Exo/SAP purification of PCR products prior to sequencing is essential to obtaining a clean read. Exo1 is intended to degrade any excess primer from the original PCR, while the SAP will dephosphorylate dNTPs from the original PCR. During the analysis, 5 µl of PCR product were purified with 0.1µl of Exo1 (20U/µl) and 1.0µl of Sap (1U/µl).

BigDye® Terminator Kit v3.1 with reduced amount of BigDye® Terminator Reagent (reduced to 1/4 for class I MHC molecules and to 1/8 for class II molecules) (Applied Biosystems) was used for labelling of the PCR product. BigDye® Terminator contains four different fluorescent dyes that label ddNTPs, which then selectively terminate chain elongation at A, C, G or T. Reactions were performed in 8µl volumes. The products were precipitated in 50µl of 96% and 70% ethanol instead of 30µl and 60µl respectively. The samples were further dissolved in Hi-Di™ Formamide (Applied Biosystems). Sequencing was performed on an ABI3730 DNA Sequencer with POP-7TM polymer and 36 cm long capillaries (Applied Biosystems). Results were analysed with SeqScape v.2.5 (Applied Biosystems).

3.3 Statistical analyses

3.3.1 Allele assignment and allele frequencies

The results were interpreted using ASSIGN-SBT software (Conexio Genomics, Applecross, Australia)(Sayer et al., 2004a). When poor or ambiguous results occurred (e.g. more than one allele was a possible match at a particular position) the procedure was repeated. If alleles or haplotype combinations were still unclear, a series of measures was taken to avoid ambiguity and among alleles chosen for further analysis, the following cases were favoured:

1) Alleles previously described in Polynesian populations;

2) Alleles with the lower number were favoured over others, rare alleles with extremely high numbers (e.g. where both B*0801 and B*0819 were possible alleles at the same position the former had been used for statistical analysis);

3) If still more than one possibility was available the most frequent allele had been chosen. Information about haplotypes and alleles discussed in this thesis, and their respective frequencies in different populations, has been compared with, and collected (in case of other populations than studied, used in numerous comparisons) from on-line allele databases such as: MHC anthropology database (dbMHC; http://www.ncbi.nlm.nih.gov/mhc/), EMBL Nucleotide Sequence Database (IMGT/HLA Database: http://www.ebi.ac.uk/imgt/hla/) and the Allele Frequencies database (http://www.allelefrequencies.net/test/). The information was additionally supplemented with data found in the articles cited.

HLA class I and class II allele frequencies were calculated by direct counting using the following formula:

$$\text{Allele frequency} = \frac{\text{sum of the individual alleles}}{2n}$$

n - the number of individuals studied; Note. In a few cases a particular DNA sample did not give a result for every locus tested.

3.3.2 Population comparisons

Two- or more-locus haplotype frequencies and linkage disequilibrium values were calculated using Arlequin 3.1 Software (Excoffier et al., 2005). Due to limited sample size and the fact that all samples represent three different populations, ready designed software was not always sufficient and erroneous. Therefore, significant amount of analysis was done manually by direct counting.

4. Results

The concentration of genomic DNA was measured before amplification. The measurement was performed using a NanoDrop® ND-1000 UV-Vis Spectrophotometer. The Table of concentrations is enclosed below. Genomic DNA was then successfully amplified in most of the samples using the "GenomiPhi DNA Amplification Kit" (Amersham Biosciences). All samples were subsequently genotyped using high resolution HLA genotyping for three class I and two class II MHC molecules. Fifteen of the 52 original samples failed to be genotyped for all the loci tested, and one sample failed to produce any results.

Sample	Conc. ng/µl	Sample	Conc. ng/µl	Sample	Conc. ng/µl	Sample	Conc. ng/µl
SM009	264.07	SM137	5.12	MA001	44.24	WS5	6.05
SM013	110.85	SM145	16.53	MA002	56.96	WS11	19.88
SM015	129.97	SM213	137.75	MA003	10.84	WS13	18.86
SM021	246.07			MA005	38.97	WS14	16.41
SM023	129.58			MA007	6.9	WS21	9.18
SM025	135.04			MA019	172.97	WS25	29.25
SM027	128.17			MA021	204.24	WS34	4.33
SM051	15.21			MA022	69.63	WS41	38.76
SM063	46.96			MA024	244.63	WS42	28.47
SM065	43.86			MA036	161.6	WS43	30.73
SM089	10.74			MA048	53.76	WS53	46.61
SM103	71.36			MA049	3.35	WS55	12.79
SM105	6.81			MA052	21.36	WS9	35.09
SM109	9.29			MA054	89.33	WS63	16.58
SM115	11.37			MA057	0.75	WS76	20.84
SM117	3.3					WS82	16.13
SM125	4.38					WS97	13.54

Table 1 List of concentrations of genomic DNA in the samples before amplification. SM: Southern Marquesas, MA: Mangareva, WS: Western Samoa.

4.1. HLA diversity and distribution

High resolution genotyping of the HLA loci revealed that most of the islanders studied carried alleles and haplotypes previously described in Polynesia, and therefore the sample set can be said to be characteristically Polynesian (Bugawan et al., 1999; Gao et al., 1997; Mack et al., 2000; Velickovic et al., 2002).

Some of the alleles observed are typical of Asian and Oceanic populations, but tend to be rare in other continents (dbMHC), e.g. B*4001, B*5502, B*5602, DRB1*1401, DQB1*0502 (Bugawan et al., 1999; Gao et al., 1997; Gao et al., 1992; Mack et al., 2000;

Maitland et al., 2004; Tracey and Carter, 2006; Velickovic et al., 2002). The relative frequencies of those alleles in this study are as follows: B*4001 – 20.4%, B*5502 – 10.2%, B*5602 – 6.12%, DRB1*1401 – 10%, DQB1*0502 – 10.7%.

The total number of potential full (extended) haplotypes was 769 (when all three populations were treated as different sample sets), obtained using Arlequin 3.1. software. Table 7 is a list of most common haplotypes; frequency ≥ 0.01. No novel alleles were observed, but several non-Polynesian and non-Asian alleles and novel allelic combinations were detected.

4.1.1. Class I alleles

HLA-A

Table 2 shows frequencies of alleles observed on the locus. Even though variation at this locus was relatively high (19 different alleles), over 90% of the allelic diversity consisted of the four allelic groups HLA-A*24 (A*2402/07/09N/11N/40N/43), HLA-A*02 (A*0201/03/06/12), HLA-A*11 (A*1101/12) and HLA-A*3401 with relative frequencies of 39.6%, 28.1%, 14.6% and 9.4% respectively. Most of the alleles were previously observed at high frequencies in Polynesia (Gao et al., 1997; Maitland et al., 2004; Severson et al., 1997). Other allele groups of HLA-A constituted only 7.3% of the variation.

Island (2n)	Southern Marquesas (n=20)	Mangareva (n=14)	Western Samoa (n=17)	Island (2n)	Southern Marquesas (n=20)	Mangareva (n=14)	Western Samoa (n=17)
A*0101		1 (0.036)		A*24	1 (0.025)	5 (0.178)	9 (0.265)
A*02	7 (0.175)	3 (0.107)	6 (0.176)	A*2402	3 (0.075)	7 (0.25)	8 (0.265)
A*0201	2 (0.05)	1 (0.036)		A*2411N	2 (0.05)		1 (0.029)
A*0203	1 (0.025)			A*2443	1 (0.025)		
A*0206	3 (0.075)	2 (0.072)	1 (0.029)	A*2501			1 (0.029)
A*0212	1 (0.025)			A*26		1 (0.036)	
A*0301	2 (0.05)			A*3301	2 (0.05)		
A*11	2 (0.05)	2 (0.072)	2 (0.059)	A*3401	2 (0.05)	1 (0.036)	6 (0.176)
A*1101	6 (0.15)	2 (0.072)		A*6811N	1 (0.025)		
A*1112		1 (0.036)					

Table 2 HLA-A Class I allele frequencies in Southern Marquesas, Mangareva and Western Samoa (frequencies are given in parentheses).

Interestingly, one of the alleles observed on Southern Marquesas (HLA-A*0212) has been used as a marker of Amerindian admixture. The highest allelic diversity on the HLA-A locus was observed in Southern Marquesas, and the lowest diversity in Western Samoa, with only 8 alleles of the most common allelic groups shown above. The most common allele was HLA-A*2402.

HLA-B

Higher levels of variation were observed on locus B. Generally, there was a significant difference between the islands, but the most frequent HLA-B alleles were: HLA-B*4001 (20%), HLA-B*4010 (12%), HLA-B*5502 (10%), HLA-B*3901 (9%) and HLA-B*5602. All the alleles were previously described in Polynesia and the relative frequencies observed here were similar to earlier observations. The commonest allele was HLA-B*4001. The most unusual HLA-B allele detected here was HLA-B*3905. This allele, previously reported in Polynesia (Gao et al., 1997; Lie et al., 2007; Maitland et al., 2004), is often used as proof of Amerindian admixture, and in the present study it was detected in Mangareva (1 donor) and Southern Marquesas (2 donors).

Island (2n)	Southern Marquesas (n=20)	Mangareva (n=14)	Western Samoa (n=17)	Island (2n)	Southern Marquesas (n=20)	Mangareva (n=14)	Western Samoa (n=17)
B*07	1 (0.025)			B*3503	1 (0.025)		
B*0702	1 (0.025)			B*3901	5 (0.125)	4 (0.143)	2 (0.059)
B*0801	1 (0.025)			B*3905	2 (0.05)	1 (0.036)	
B*1402	2 (0.05)			B*4001	2 (0.05)	4 (0.143)	14 (0.412)
B*1501	2 (0.05)	4 (0.143)	1 (0.029)	B*4002	2 (0.05)		3 (0.088)
B*1502	1 (0.025)			B*4010	6 (0.15)	4 (0.143)	2 (0.059)
B*1506			2 (0.059)	B*4056			1 (0.029)
B*1517		1 (0.036)		B*4801	4 (0.1)	1 (0.036)	3 (0.088)
B*1801		1 (0.036)		B*5502	5 (0.125)	4 (0.143)	1 (0.029)
B*2704			1 (0.029)	B*5601	1 (0.025)		
B*2705	1 (0.025)			B*5602		2 (0.071)	4 (0.118)
B*35	1 (0.025)						

Table 3 HLA-B Class I allele frequencies in Southern Marquesas, Mangareva and Western Samoa (frequencies are given in parentheses).

HLA-Cw

Out of all loci tested, the greatest diversity was observed on HLA-Cw, with more than 20 alleles. The greatest HLA-Cw diversity was in Southern Marquesas, and the lowest in Mangareva. Almost all allelic groups (9) were found on Western Samoa, Mangareva lacked HLA-Cw*14 and HLA-Cw*15, and Southern Marquesas lacked HLA-Cw*03. The commonest groups were HLA-Cw*04, HLA-Cw*03, HLA-Cw*01 and HLA-Cw*07, with relative frequencies of 20%, 18%, 18% and 16%. The most common allele was HLA-Cw*0102.

Island (2n)	Southern Marquesas (n=20)	Mangareva (n=14)	Western Samoa (n=17)	Island (2n)	Southern Marquesas (n=20)	Mangareva (n=14)	Western Samoa (n=17)
Cw*0102	5 (0.125)	7 (0.25)	6 (0.176)	Cw*0709	1 (0.025)		
Cw*0202	1 (0.025)			Cw*0727		2 (0.071)	
Cw*0303		2 (0.071)	1 (0.029)	Cw*0801	5 (0.125)	1 (0.036)	2 (0.059)
Cw*0304		2 (0.071)	12 (0.323)	Cw*0802	2 (0.05)		
Cw*0309			1 (0.029)	Cw*0810			1 (0.029)
Cw*0311		1 (0.036)		Cw*1202	1 (0.025)		1 (0.029)
Cw*04			1 (0.029)	Cw*1203	1 (0.025)	2 (0.071)	1 (0.029)
Cw*0401	4 (0.1)	5 (0.178)		Cw*1402	1 (0.025)		
Cw*0403	4 (0.1)	4 (0.143)	3 (0.088)	Cw*15			1 (0.029)
Cw*07	1 (0.025)			Cw*1502	1 (0.025)		1 (0.029)
Cw*0701	1 (0.025)			Cw*1503			1 (0.029)
Cw*0702	8 (0.2)	2 (0.071)	2 (0.059)	Cw*1507	2 (0.05)		

Table 4 HLA-Cw Class I allele frequencies in Southern Marquesas, Mangareva and Western Samoa (frequencies are given in parentheses).

4.1.2. Class II alleles

HLA-DRB1

The three most common HLA-DRB1 allelic groups were HLA-DRB1*09 (23%), HLA-DRB1*14 (19%) and HLA-DRB1*04 (14%). The relative frequency differed between islands, but the most common alleles were HLA-DRB1*0901, HLA-DRB1*0403 and HLA-DRB1*1401. Interestingly, one allele (HLA-DRB1*070101) found at high frequency (14%) in Mangareva was absent in the other two islands. The particular allele is typical of Caucasians.

Island (2n)	Southern Marquesas (n=20)	Mangareva (n=14)	Western Samoa (n=17)	Island (2n)	Southern Marquesas (n=20)	Mangareva (n=14)	Western Samoa (n=17)
DRB1*0102	1 (0.025)			DRB1*1202	1 (0.025)		
DRB1*0301	2 (0.05)			DRB1*1203		1 (0.036)	
DRB1*0403	4 (0.1)	6 (0.214)	3 (0.088)	DRB1*1205		1 (0.036)	
DRB1*0405			1 (0.029)	DRB1*1206			1 (0.029)
DRB1*0701		4 (0.143)		DRB1*1301		1 (0.036)	
DRB1*0802	1 (0.025)			DRB1*1302		1 (0.036)	
DRB1*0803	2 (0.05)	3 (0.107)	5 (0.147)	DRB1*1401	6 (0.15)	2 (0.071)	1 (0.029)
DRB1*0901	6 (0.15)	3 (0.107)	15 (0.441)	DRB1*1402	1 (0.025)		
DRB1*11	2 (0.05)	1 (0.036)	2 (0.059)	DRB1*1407	1 (0.025)		
DRB1*1101	2 (0.05)	3 (0.107)	1 (0.029)	DRB1*1408	2 (0.05)		3 (0.088)
DRB1*12	3 (0.075)			DRB1*1501	1 (0.025)		1 (0.029)
DRB1*1201	1 (0.025)	2 (0.071)					

Table 5 HLA-DRB1 Class II allele frequencies in Southern Marquesas, Mangareva and Western Samoa (frequencies are given in parentheses).

HLA-DQB1

This exhibited the lowest diversity, with only 11 alleles observed, within 5 allelic groups. The commonest alleles were HLA-DQB1*0301 and HLA-DQB1*0303.

Island (2n)	Southern Marquesas (n=20)	Mangareva (n=14)	Western Samoa (n=17)	Island (2n)	Southern Marquesas (n=20)	Mangareva (n=14)	Western Samoa (n=17)
DQB1*0201	1 (0.025)	4 (0.143)		DQB1*0402	2 (0.05)		
DQB1*0301	9 (0.225)	5 (0.178)	3 (0.088)	DQB1*0501	3 (0.075)		
DQB1*0302	5 (0.125)	6 (0.214)	1 (0.029)	DQB1*0502	7 (0.175)	2 (0.071)	
DQB1*0303	6 (0.15)	3 (0.107)	8 (0.235)	DQB1*0503	5 (0.125)	2 (0.071)	2 (0.059)
DQB1*0309	1 (0.025)			DQB1*0601	1 (0.025)		6 (0.176)
DQB1*0401			2 (0.059)				

Table 6 HLA-DQB1 Class II allele frequencies in Southern Marquesas, Mangareva and Western Samoa (frequencies are given in parentheses).
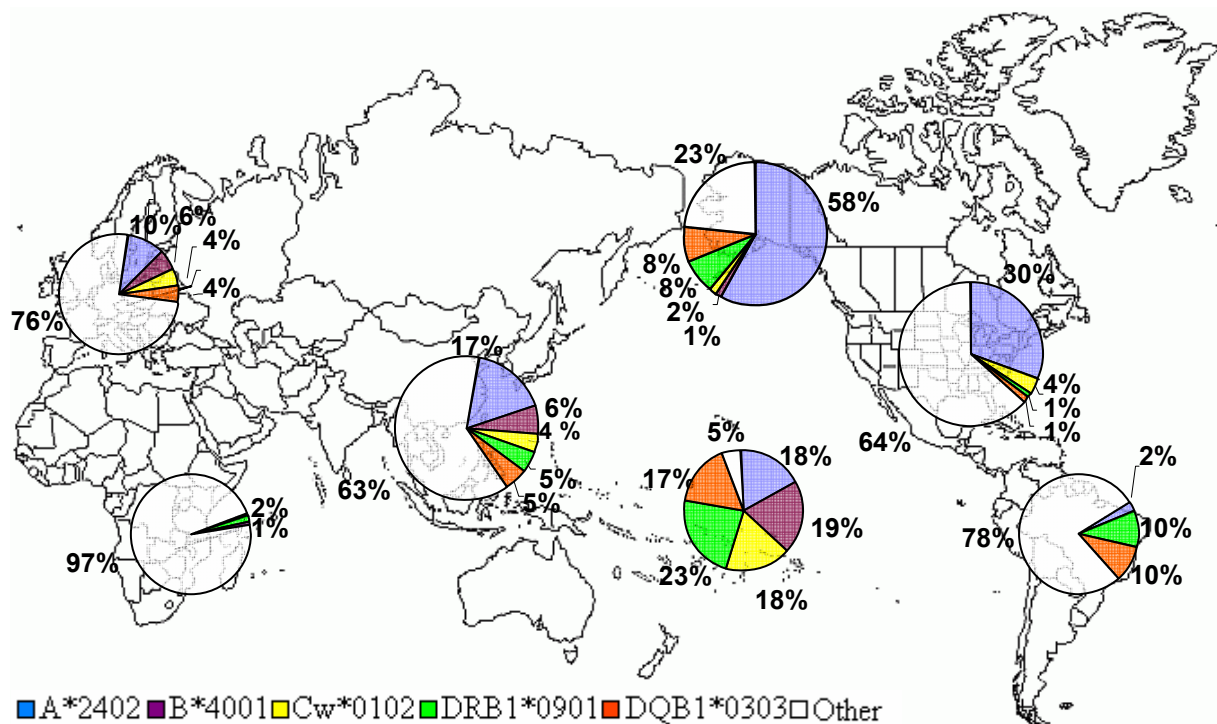

Figure 3 Most common alleles found on each locus in Eastern Polynesia and their distribution in six other populations throughout the world (Malay from Singapore, Shona from Zimbabwe, Central European represented by the Finn and Czech, G. Kaiowa from Brazil, Canoncito from New Mexico and Yupik from Alaska). The percentage values show relative frequency of these alleles in the populations. The values smaller than 1% have not been included. (The map was from http://worldatlas.com). The figure also depicts relatively low (reduced) heterozygosity among Pacific islanders.

In conclusion, the presence of Asian elements such as DRB1*0405 and 1201 among the class II molecules observed, could point to a Polynesian origin further north than South East Asia. Moreover, the high frequency of DRB1*0403 suggests a Melanesian contribution to the Polynesian gene pool, since this allele is common in costal Melanesians (Gao et al., 1992). The distribution of the allele becomes more prominent in Eastern Polynesia and reaches its highest frequency peak in the island of Mangareva and lowest in Western Samoa. Such distribution could be the result of 'Express train' model of colonization followed by subsequent founder effects and population bottlenecks. *'Unique to Polynesians among Oceanic populations was the presence of DRB1*0802, described previously only in Amerindians; other Polynesian features were the relatively high frequencies of DRB1*0901, a common allele in East Asians and the novel DRB1*1408'*(Gao et al., 1992). Interestingly, in this study the DRB1*1408 allele described by Gao et al. had the highest frequency on Samoa and Southern Marquesas but was completely missing from Mangareva.

HLA-A	HLA-B	HLA-Cw	HLA-DRB1	HLA-DQB1	Frequency	Possible origin
A*02	B*4001	Cw*0304	DRB1*0901	DQB1*0303	0.0392	Polynesian
A*24	B*4001	Cw*0304	DRB1*0901	DQB1*0303	0.0196	Polynesian
A*24	B*3901	Cw*0303	DRB1*0701	DQB1*0502	0.0196	SEAsian/Oceanic
A*1101	B*4010	Cw*0403	DRB1*0403	DQB1*0302	0.0392	Polynesian

Table 7 Most frequent haplotypes – all of Polynesian origin (Arlequin 3.1)

The population comparison analysis was performed as difference/distance analysis between the populations of Samoa, Mangareva and Marquesas. Later, the samples were grouped to form one sample set, called the Eastern Polynesian sample, and compared to other populations from different continents, using data from dbMHC (http://www.ncbi.nlm.nih.gov /mhc/) to assess the phylogenetic relationships between the populations and infer the origins of Eastern Polynesians.

Population pairwise Fst value analysis showed that Southern Marquesas and Mangareva are most similar (Fst=0.00513) and Southern Marquesas and Western Samoa are most different populations in this study (Fst=0.06019). This could be due to genetic drift within French Polynesia and admixture within the archipelago, as well as geographical distance from Samoa.
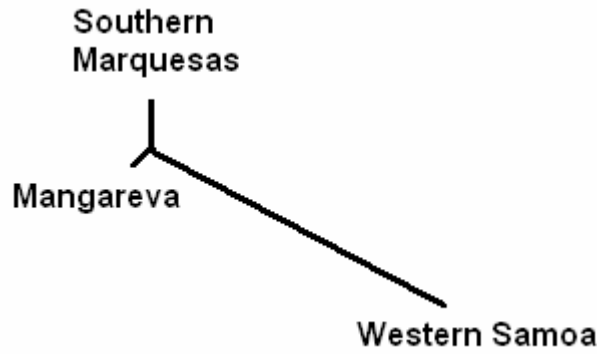
Figure 4 Neighbor Joining tree representing the three populations and the relationships between them based on pairwise Fst values.

For all three islands, homozygosity was a little lower than expected, but significant deviations of F were only observed in the class II loci (Table 3). Interestingly, the reduced heterozygosity of DQB1 in Samoa has been reported previously (Mack et al., 2000; Velickovic et al., 2002) and is said to reflect balancing selection in this population.

Locus	Population	Observed heterozygosity	Expected heterozygosity	P-value	Significance
HLA-A*	Marquesas	0.77778	0.92540	0.00601	n.s.
	Mangareva	0.76923	0.88615	0.12225	n.s.
	Samoa	0.76471	0.83066	0.09594	n.s.
HLA-B*	Marquesas	0.73684	0.94168	0.03345	n.s.
	Mangareva	0.76923	0.90462	0.21045	n.s.
	Samoa	0.70588	0.82888	0.10311	n.s.
HLA-Cw*	Marquesas	0.78947	0.91181	0.03581	n.s.
	Mangareva	0.78571	0.88889	0.71014	n.s.
	Samoa	0.82353	0.86096	0.19977	n.s.
HLA-DRB1*	Marquesas	0.78947	0.94168	0.04312	n.s.
	Mangareva	0.78571	0.91534	0.00043	<0.05
	Samoa	0.76471	0.78253	0.19903	n.s.
HLA-DQB1*	Marquesas	0.80000	0.87692	0.80204	n.s.
	Mangareva	0.45455	0.84416	0.00023	<0.05
	Samoa	0.45455	0.79221	0.00194	<0.05

Table 8 Heterozygosity (F) values for all the loci. Most values were not significant (n.s.) due to small sample size.

	HLA-A	HLA-B	HLA-Cw	HLA-DRB1	HLA-DQB1	Possible origin	Observed in
1	26	3905	0401	11	0503	Amerindian/ Mixed	Mangareva
2	0201	3905	1203	1407	0402	Amerindian/Unknown	S. Marquesas
3	0212	4010	1402	0901	0303	Mixed	S. Marquesas
4	1101	3905	0702	0802	0402	Hispanic /Amerindian	S. Marquesas

Table 9 Haplotypes containing Amerindian alleles (Arlequin 3.1). The haplotypes were observed on two of the islands studied and each of them occurred only once.

5. Discussion

5.1. Amerindian alleles detected in Eastern Polynesia

In this analysis of HLA variation in Eastern Polynesia, the most frequent haplotypes detected were clearly Polynesian. However, four haplotypes contained the Amerindian alleles B*3905 and A* 0212 (shown in Table 9). Both these alleles have been shown to be of Amerindian ancestry (Belich et al., 1992; Layrisse et al., 2001; Martinez-Arends et al., 1998) and are rare in Polynesian and non-Amerindian populations (Bugawan et al., 1999; Gao et al., 1997). The detection of these alleles, thought to be of South American origin (Belich et al., 1992; Watkins et al., 1992), in the Easter Island population, has been suggested as proof of prehistoric migrations into the Pacific of South American navigators (Lie et al., 2007). It is feasible that their presence in Mangareva and Southern Marquesas is also indicative of South American contacts.

The haplotypic context of individual alleles can yield important information on their origins, as HLA loci are closely linked and tend to be inherited as a haplotype. Moreover, linkage disequilibrium between loci results in recombination at different frequencies within the HLA encoding region. It can be assumed that recombination happens over time and the haplotypes will become gradually different from the original parental lineages.

Following this line of reasoning, I searched the available published data for the typical haplotypic surroundings of the alleles detected in this study. The search for the combination of B*3905-Cw*0401, as seen on haplotype No. 1 (Table 9), showed that 8 out of 10 hits in the dbMHC anthropology database are represented by South American Amerindians (Guarani-Kaiowa). The combination of A*0201 and B*3905, as seen on haplotype No. 2, is also very common among Amerindians, and was detected in 28% of the hits for B*3905 in the database, mostly in people of Amerindian origin, while just one single hit for B*3905-Cw*1203 was detected in a Hispanic population. A search for the two-locus combination B*3905 and Cw*0702 as seen on haplotype No. 4 resulted in 52 hits of mainly Amerindian origin (38 individuals), but the exact combination of the three observed loci A*1101-B*3905-Cw*0702 was only seen in two subjects, both USA Hispanics (Cao et al., 2001). The 4-locus haplotype: B*3905-Cw0702-*DRB1*0802-DQB1*0402 (haplotype No. 4) has only been detected in Amerindians previously. And the class II haplotype DRB1*0802-DQB1*0402, occurring on the same extended haplotype, was observed in a Mexican Mestizo population,

and both North and South American populations (Garcia-Ortiz et al., 2006; Petzl-Erler et al., 1997).

Therefore, it is highly probable that the Amerindian alleles observed in this study were brought into Polynesia during prehistory, as they occur on different haplotypic backgrounds. Moreover, the alleles are usually in novel combinations, suggesting that significant time has elapsed since their introduction into the genetic pool of the islanders, as for example in the case of haplotype No. 3.

If the Arlequin analysis is disregarded (as the following haplotypes were not listed among most frequent extended haplotypes), the two B*3905 alleles in Southern Marquesas sample could be interpreted as belonging to the haplotypes B*3905-Cw*0702-DQB1*0402. This combination has been shown to occur at a frequency of 5.6% in the South American tribes Guarani-Nandewa and Guarania-Kaiowa (dbMHC Anthropology Database).

Allele DRB1*1402 found in Southern Marquesas gave 357 hits on dbMHC, of which most were native inhabitants of both Americas and a few Aboriginal Australians. Only four individuals were Africans and one of mixed origin. The allelic combination on which this allele occurred (DRB1*1402-DQB1*0301) was previously detected in Amerindian populations.

Another Amerindian allele found in the Marquesas and Samoa was B*4801 (Petzl-Erler et al., 1997), previously shown to occur at significant frequencies in Samoa (Severson et al., 1997). However, it has been suggested that the HLA-B*48 alleles probably reflect common Asian ancestry rather than direct contact between Polynesia and South America (Gao et al., 1992).

Regarding the A*0206-B*2705-C*0202 allele combination observed once in Southern Marquesas, while most of the alleles have wide geographical distribution and cannot be called population-specific, the combination itself has only been observed in indigenous inhabitants of Northern America.

In conclusion, numerous alleles and allele combinations detected during this analysis point to possible prehistoric trans-Pacific contacts between Polynesia and the Americas.

5.2. European alleles detected in Eastern Polynesia

As mentioned above, some of the alleles detected in this study were previously described, and due to their population specificity can be used as indicators of possible ancestry. Several alleles detected in this study can be used as European markers. Comparison of the presumed "European" alleles with the available databases revealed the following

results: Cw*0802 allele gave many hits in populations described as Hispanic, European and Sub-Saharan Africans. Only three such alleles were found in Oceania and Australia. In my sample set the allele occurs on the following haplotypes from the Southern Marquesas:

	HLA-A	HLA-B	HLA-Cw	HLA-DRB1	HLA-DQB1	Possible origin	Observed in
1	24	1402	0802	0803	0503	Hispanic, European /Aboriginal Australian, PNG Highlander	S. Marquesas
2	3301	4002	0802	1408	0501	European/Mixed	S. Marquesas
3	11	1501	0304	0701	0303	European/Mixed	Mangareva
4	0201	1517	0304	1302	0201	European	Mangareva

Table 10 Haplotypes containing "Caucasoid" alleles (Arlequin 3.1)

It is worth mentioning that of 256 hits for B*1402 in dbMHC database, only one was a Filipino, and all other donors mostly of European and African origin (with a predominance of European alleles). It is therefore interesting that the first haplotype in the table above appears a hybrid of European and Aboriginal Australian haplotypes (Mack et al., 2000; Velickovic et al., 2002).

The European DRB1*070101 allele was found in four individuals of Mangareva. The search for the allele in various databases resulted in finds of mostly that origin. Almost all the allelic combinations of this allele in Mangareva have been described in the Irish. For example, haplotype A*0101-B*1501-Cw*0304-DRB1*0701 obtained during direct observation analysis was found to be a perfect match among the Irish. While this could be proof of European admixture, it should be noted that DRB1*0701 (the shortened numeric version of the allele name) produced many more hits, of many populations, possibly as a result of the different types of methods used for determining alleles. The presence of the allele might be the result of African, Amerindian or Asian origin, or of admixture, since it was previously described in Polynesian populations (Mack et al., 2000).
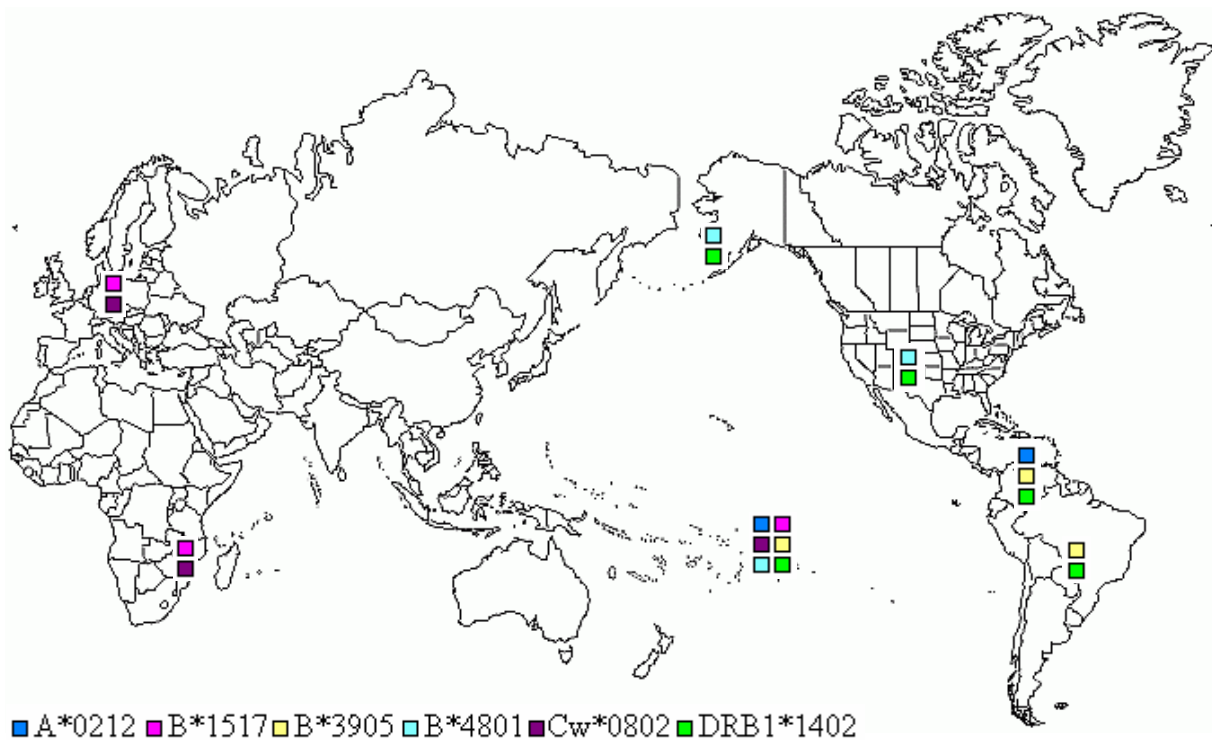
Figure 5 The 'rarest' alleles found in Eastern Polynesia and their distribution in six other populations throughout the world (Malay from Singapore (none of the alleles observed), Shona from Zimbabwe, Central European represented by the Finn and Czech, G. Kaiowa from Brazil, Yucpa from Venezuela (Layrisse et al., 2001), Canoncito from New Mexico and Yupik from Alaska). The alleles tend to be population-specific and therefore the boxes also represent their probable origins. Relative frequencies of the alleles are shown in Figure 6. (The map was from http://worldatlas.com).
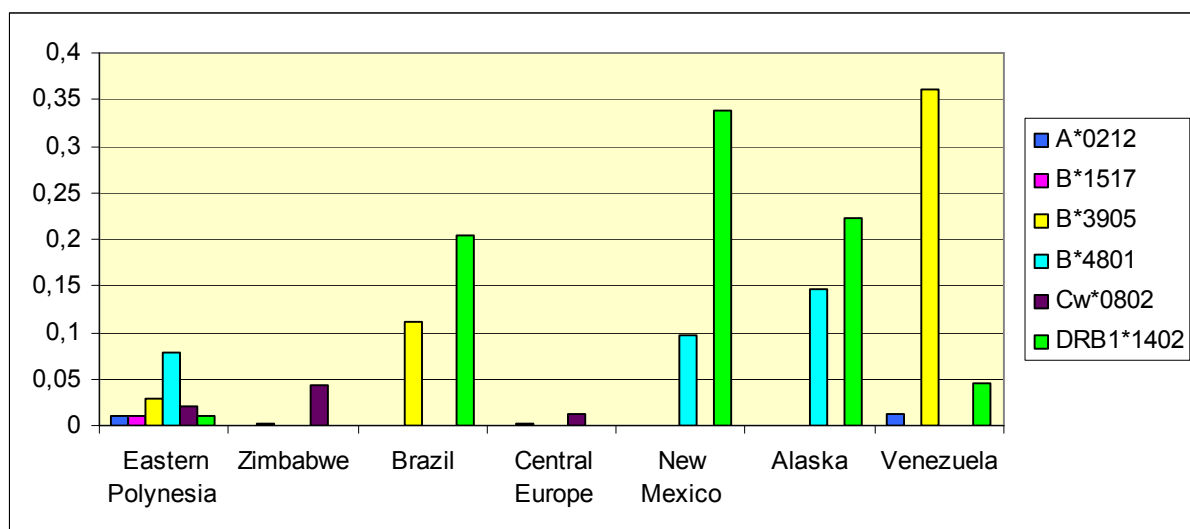

Figure 6 The table shows relative frequencies of 'rare' alleles found in Polynesia and their distribution in different populations. As expected, most of the alleles were previously described in Native American populations and possibly arose as a result of trans-Pacific contacts between the Americas and Polynesia.

The allele B*1517 could probably be more informative since it is found mainly in European and African populations, and only rarely in Asia. Notably, in the Pacific it has only been observed in Australian whites (Gao et al., 1997). Allele B*1517 was assigned by

Arlequin analysis to haplotype A*0201-B*1517-Cw*0304-DRB1*1302-DQB1*0201. The search for haplotypes consisting of the A/B/Cw loci resulted in two exact matches, one of Amerindian and one of European origin. Of the two, the latter seems to be more probable source since it also carried the DRB1*1302 allele (of mainly African, European and Asian ancestry). Moreover, the combination of DRB1*1302-DQB1*0201 has only been described in African and European populations (with one donor from Asia). It is therefore likely that this particular allele might be result of recent European admixture.

It is worth mentioning that the above analysis was mainly based on the haplotypes generated by Arlequin 3.1. While precautions were taken to avoid ambiguities, the program cannot be fully trusted since it is designed for larger sample sets. Moreover, the analysis could be biased as the program reads 0 (no allele) as 0-allele (allele called '0') and creates haplotypes on the basis of non-existing alleles. The program was used in order to obtain recurrent results, but additional manual analysis was performed. Manual analysis, though subjective, resulted in a much more exhaustive list of most frequent haplotypes, most of which were previously described in Polynesia in other studies. The haplotypes are not mutually exclusive, since Arlequin displays only the most frequent haplotypes, but the possible haplotypes greatly exceed listed results. It is probable that the extended list of haplotypes proposed by Arlequin in fact contains haplotypes identical to those detected by direct analysis.

HLA-A	HLA-B	HLA-Cw	HLA-DRB1	HLA-DQB1	Possible origin
A*24	B*4001	Cw*0304	DRB1*090102	DQB1*030302	Polynesian
A*1101	B*4010	Cw*0403	DRB1*0403	DQB1*0302	Polynesian
A*24	B*3901	Cw*0702	DRB1*1401	DQB1*050201	Polynesian
A*02	B*5502	Cw*0102	DRB1*12	DQB1*0301	Polynesian
A*3401	B*4002	Cw*15	DRB1*1408	DQB1*050301	Australia/ Oceania
A*24	B*4001	Cw*0401	DRB1*1101	DQB1*0301	Polynesian
A*11	B*5602	Cw*0102	DRB1*090102	DQB1*030302	Polynesian
A*0201	B*1501	Cw*0304	DRB1*070101	DQB1*0201	European

Table 11 Most frequent haplotypes obtained by direct counting.

Phylogenetic analysis was performed using the PHYLIP package available online (Felsenstein, 2004). Gene frequency based distance matrix was introduced to the program and a consensus tree was plotted to establish the relationship between the populations in this study and other global populations. A Neighbor-Joining tree was generated using PHYLIP. As was the case with the analysis of pairwise Fst's, Western Samoa and Mangareva tend to cluster more closely than Western Samoa and Southern Marquesas. But the populations of Southern

Marquesas and Mangareva were most closely related, not surprising in view of geographical proximity and the fact that Mangareva was settled from Marquesas. The difference between Southern Marquesas and Western Samoa is probably due to the dramatic population collapse of 19th century, following European contacts. Europeans brought disease (measles, venereal diseases and influenza), war and forced labour, and a great famine occurred in 1804. The population of Marquesas decreased from around 100,000 in 1773 to just 4279 in 1897 (Campbell, 1989). Such events must have had profound consequences for the patterns of genetic variation observed in present-day populations.
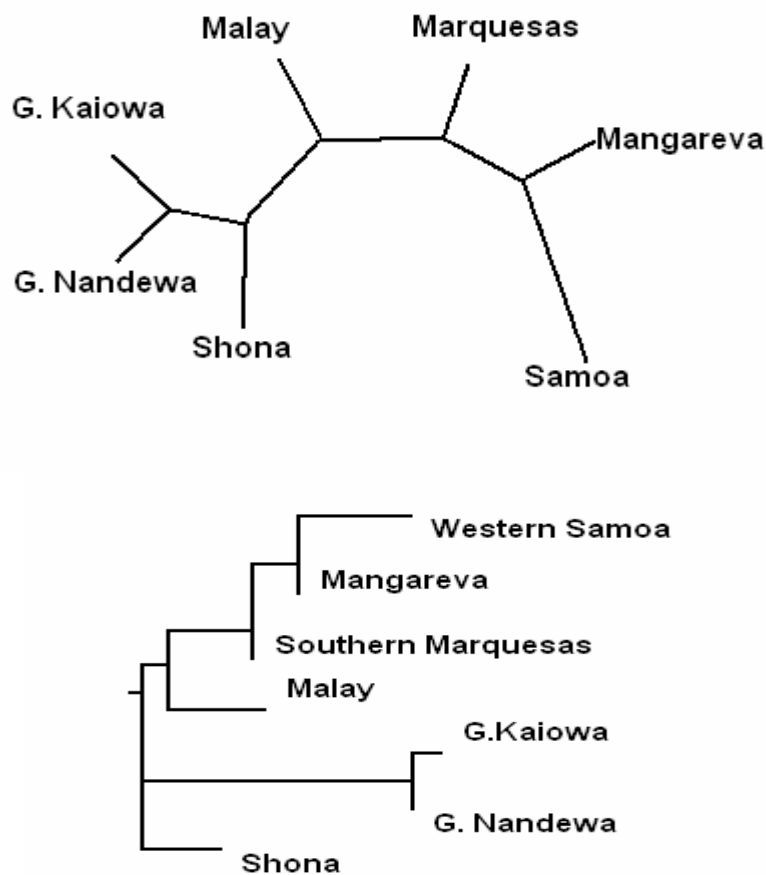

Figure 7 Neighbor-Joining (A) and Consensus (B) trees representing phylogenetic relationships between seven populations tested on the basis of 5-loci polymorphisms. The trees were constructed using pairwise differences between populations calculated from observed allele frequencies.

As seen in Figure 7, the topography of phylogenetic trees suggests that the Eastern Polynesian cluster is most closely related to Malays and distinctly separated from both Amerindian populations and the African one. However, the trees should be interpreted with

caution. HLA alleles, though very informative in population studies, are subject to selective pressures which can influence the relative frequencies of alleles and consequently distort the overall picture of population relationships. For example, positive selective pressure for A*2402 in PNG highlands may be due to resistance to disease pathogens.

The final analysis used Arlequin 3.1 to test the significance of the pairwise differences between 9 Oceanic populations for which data was available on dbMHC. Unfortunately, most of the data available consisted of just three-locus HLA polymorphisms (HLA-A, HLA-DRB1 and HLA-DQB1). Nevertheless, the three populations of this study were compared to Malay, Filipino, Moluccan, East Timorese, PNG Highlander and PNG Lowlander populations, to test wider relationships, albeit on a narrower set of loci. The results showed no significant differences between the populations, with the exception of PNG Highlanders and PNG Lowlanders, who not only differed from other populations but were significantly different to one another. The only other pair of significantly different populations was the Moluccan and Filipino. This suggests that Eastern Polynesians of this study are not significantly different from the populations of Indonesia and the Philippines but differ from the New Guineans. This is similar to the results obtained using mtDNA analyses, and is consistent with an 'Express train model' of the colonization of the Pacific.

6. Conclusion

The three human DNA samples from Eastern Polynesia exhibit typical Polynesian HLA alleles. But other, rare alleles, point to possible past admixture with Amerindian genes in Southern Marquesas and Mangareva, and European genes in Mangareva. The analysis of the haplotypes on which those rare alleles occurred show that the Amerindian alleles usually occur in novel haplotypic combinations together with Polynesian alleles, and therefore predate the European elements, which generally appear in European haplotypic combinations. In other words, the Amerindian alleles appear to have been introduced into the Polynesian gene pool much earlier than the European alleles. Lack of genealogical information on the individuals included in this study prevented the estimation of the relative time of admixture. Therefore, it is not possible to conclude definitively whether the admixture is the result of the 19th century Peruvian slave raids, as was suggested for the island of Rapa (Hurles et al., 2003), or due to prehistoric human migrations as suggested for the Amerindian alleles detected in Easter Island (Lie et al., 2007).

This study confirms the widely held view that present-day Polynesians originated in Asia, and supports the 'Express train model' of the colonization of the Pacific. But it does not rule out the possibility of trans-Pacific contacts in prehistory. This study should be complemented by additional analyses of mtDNA and Y chromosomes to shed light on the origin of the maternal and paternal lineages in East Polynesia.

Appendix

Table 1. Amplification primers

LOCUS	PRIMER	ANNEALING POSITION	SEQUENCE
HLA-A [1]	HLA-A 5APM13F	5UTR	TGT AAA ACG ACG GCC AGT TCT CCC CAG ACG CCG AGG ATG GCC
	3AE4.658M13R	Exon 4	CAG GAA ACA GCT ATG ACC AGT CCT GGG TCT GGT CCT CCC CAT
HLA-B [2]	BIN1-CGM13F	Intron 1	TGT AAA ACG ACG GCC AGT CGG GGG CGC AGG ACC CGG
	BIN1-TAM13F	Intron 1	TGT AAA ACG ACG GCC AGT GGC GGG GGC GCA GGA CCT GA
	BIN37DM13R	Intron 3	CAG GAA ACA GCT ATG ACC AGG CCA TCC CSG SCG AYC TAT
HLA-C [1]	HLACM13F	Intron 1	TGT AAA ACG ACG GCC AGT AGC GAG GKG CCC GCC CGG CGA
	HLAC+15M13R	Intron 3	CAG GAA ACA GCT ATG ACC GGA GAT RGG GAA GGC TCC CCA CT
HLA-DRB1 [3]	DRB1-52.1		CAG GAA ACA GCT ATG ACC CCC ACA GCA CGT TTC TTG GAG TAC TCT A
	DRB1-01		CAG GAA ACA GCT ATG ACC TGA GAC GCA CGT TTC TTG TGG CAG CTT AAG TT
	DRB1-04		CAG GAA ACA GCT ATG ACC TGA GAC GCA CGT TTC TTG GAG CAG GTT AAA C
	DRB1-09		CAG GAA ACA GCT ATG ACC TGA CCA GCA CGT TTC TTG AAG CAG GAT AAG TT
	DRB1-10		CAG GAA ACA GCT ATG ACC TGA AGA CCA CGT TTC TTG GAG GAG G
	DRB1-15		CAG GAA ACA GCT ATG ACC TGA GAC TCA CGT TTC CTG TGG CAG CCT AAG A
	DRB1-07		CAG GAA ACA GCT ATG ACC TGA GAC TCA CGT TTC CTG TGG CAG GGT AAG TAT A
	REVERSE		TGT AAA ACG ACG GCC AGT GCT YAC CTC GCC KCT GCA C
HLA-DQB1 [1]	DQB1PF2_M13F	Intron 1	TGT AAA ACG ACG GCC AGT CCY CGC AGA GGA TTT CGT G
	DQB5/6_M13R	Exon 2	CAG GAA ACA GCT ATG ACC CTC TCC TCT GCA RGA TCC C
	DQB2/3/4_M13R	Exon 2	CAG GAA ACA GCT ATG ACC CTC GCC GCT GCA AGG TCG T

[1] B.A. Lie, personal communication; Primer sequences based on unpublished results and IHWG strategies (http://www.ihwg.org/)
[2] Primers designed upon dimorphic primers by Cereb et Young and IHWG strategies (http://www.ihwg.org/)
[3] (Sayer et al., 2004b)

Table 2. Sequencing primers

LOCUS	PRIMER	ANNEALING POSITION	SEQUENCE
HLA-A	ASEQ3	Intron 3	TCG GAC CCG GAG ACT GTG
	ASEQ5	Intron 1	GTT TCA TTT TCA GTT TAG GCC A
	3Aln3-66	Intron 3	TGT TGG TCC CAA TTG TCT CCC CTC
	5Aln1-46	Intron 1	GAA ACS GCC TCT GYG GGG AGA AGC AA
HLA-B	E2	Intron 2	AAC TGA AAA TGA AAC CGG GT
	E3	Intron 2	ACC CGG TTT CAT TTT CAG TT
	M13F	M13 tail	TGT AAA ACG ACG GCC AGT
	M13R	M13 tail	CAG GAA ACA GCT ATG ACC
HLA-C	M13F		TGT AAA ACG ACG GCC AGT
	M13R		CAG GAA ACA GCT ATG ACC
	HLACSEQR123	Intron 2	GGA GRC GTG ACC TGC GCC CCR GG
	CF1CSEQ	Intron 2	CGG GGG GGG GGC CAG
HLA-DRB1	M13F		TGT AAA ACG ACG GCC AGT
	M13R		CAG GAA ACA GCT ATG ACC
HLA-DQB1	M13F		TGT AAA ACG ACG GCC AGT
	M13R		CAG GAA ACA GCT ATG ACC

Table 3. Reagents

Product	Locus	Supplier	Product Number
GenomiPhi™ Amplification Kit	All	Amersham Biosciences,	25-6600-01
50bp Gene Ruler	HLA-DRB1,-DQB1	Fermentas, Ontario L7N 3N4, Canada	SM0372
1kb Gene Ruler	HLA-A,-C, -B	Fermentas, Ontario L7N 3N4, Canada	SM0311
SAP (Shrimp Alkaline Phosphatase)	All	Roche Diagnostics Norge AS, 0607 Oslo, NO	1758250
EXO1 (Exonuclease I)	All	Sigma-Aldrich, St. Louis, MO 63103, USA	M0293L
Agarose	All	Sigma-Aldrich, St. Louis, MO 63103, USA	A9311
Ethidium Bromide	All	Mercury, Pretech Instruments KB, 191 44 Sollentuna, Sweden	5450
BigDye® Terminator Kit v3.1	All	Applied Biosystems, Foster City, CA 94404, USA	4337456
Ethanol 100%	All	Sigma-Aldrich, St. Louis, MO 63103, USA	32205
Hi-Di™ Formamide	All	Applied Biosystems, Foster City, CA 94404, USA	4311320
Platinum® Taq DNA Polymerase	HLA-C, -B	Invitrogen Corporation, Carlsbad, CA 92008, USA	10966-034/ 11509-015
Platinum® $MgSO_4$		Invitrogen Corporation, Carlsbad, CA 92008, USA	10966-034/ 11509-015
Platinum® 10X PCR_x Amplification Buffer	HLA-C, -B	Invitrogen Corporation, Carlsbad, CA 92008, USA 11509-015	11509-015
Platinum® 10X PCR_x Enhancer Solution	HLA-C, -B	Invitrogen Corporation, Carlsbad, CA 92008, USA	11509-015
dNTPs	All	Bioline Ltd., London NW2 6EW, GB DNA Technology	BIO-39026
10x NH_4-based Reaction Buffer	HLA-DRB1	Bioline Ltd., London NW2 6EW, GB	BIO-21040
BIOTAQ™	HLA-DRB1	Bioline Ltd., London NW2 6EW, GB	BIO-21040
MgCl2 Bioline	HLA-DRB1	Bioline Ltd., London NW2 6EW, GB	BIO-21040
10x buffer	HLA-DQB1	Promega U.S., WI 53711, USA	M1668
Taq DNA Polymerase	HLA-DQB1	Promega U.S., WI 53711, USA	M1668
$MgCl_2$	HLA-DQB1	Promega U.S., WI 53711, USA (Nerlines Meszansky)	M1668
PfuTurbo® DNA Polymerase	HLA-DQB1	Stratagene, La Jolla, CA 92037 (MedProbe)	600252
Bovine Serum Albumin (BSA)	HLA-DQB1	New England Biolabs Ipswich, MA 01938-2723	B9001S
Primers	All	MedProbe, N-0131 Oslo, NO	Custom made

Acknowledgments

References

Anderson, A. (1991). The chronology of colonization in New Zealand. Antiquity *65*, 767-795.

Belich, M.P., Madrigal, J.A., Hildebrand, W.H., Zemmour, J., Williams, R.C., Luz, R., Petzl-Erler, M.L., and Parham, P. (1992). Unusual HLA-B alleles in two tribes of Brazilian Indians. Nature *357*, 326-329.

Bellwood, P. (1978). Man's Conquest of the Pacific. The Prehistory of Southeast Asia and Oceania (Auckland, William Collins Publishers Ltd.).

Bellwood, P. (1987). The Polynesians. Prehistory of an island people (London, Thames and Hudson Ltd.).

Bellwood, P. (1991). The Austronesian Dispersal and the Origin of Languages. Scien Am 70-75.

Borrell, B. (2007). Drifters could explain sweet-potato travel; An unsteered ship may have delivered crop to Polynesia. Nature.

Browning, M., and McMichael, A. (1996). HLA and MHC: genes, molecules and function (Oxford, Bios Scientific Publ.).

Brumester, G.R., and Pezzutto, A. (2003). Color Atlas of Immunology (Stuttgard, New York, Thieme).

Bugawan, T.L., Mack, S.J., Stoneking, M., Saha, M., Beck, H.P., and Erlich, H.A. (1999). HLA class I allele distributions in six Pacific/Asian populations: evidence of selection at the HLA-A locus. Tissue Antigens *53*, 311-319.

Campbell, I.C. (1989). A History of the Pacific Islands (Berkeley, Los Angeles, University of California Press).

Cao, K., Hollenbach, J., Shi, X., Shi, W., Chopek, M., and Fernandez-Vina, M.A. (2001). Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations. Hum Immunol *62*, 1009-1030.

Cavalli-Sforza L. L., M.P., Pizza A. (1996). The History and Geography of Human Genes (New Jersey Princeton University Press).

Diamond, J.M. (1988). Express train to Polynesia. Nature *336*, 307-308.

Diamond, J.M. (2000). Taiwan's gift to the world. Nature *403*, 709-710.

Dunn, M., Terrill, A., Reesink, G., Foley, R.A., and Levinson, S.C. (2005). Structural phylogenetics and the reconstruction of ancient language history. Science *309*, 2072-2075.

Excoffier, Laval, L.G., and Schneider, S. (2005). Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online *1*, 47-50.

Felsenstein, J. (2004). PHYLIP (Phylogeny Inference Package) version 3.6. (Seattle, Department of Genome Sciences, University of Washington), pp. Distributed by the author.

Fischer, S.R. (2002). A History of the Pacific Islands (New York, Palgrave).

Gao, X., Lester, S., Boettcher, B., and McCluskey, J. (1997). Diversity of HLA genes in populations of Australia and the Pacific. Paper presented at: HLA Genetic diversity of HLA Functional and Medical Implication (EDK: Medical and Scientific International Publisher).

Gao, X., Zimmet, P., and Serjeantson, S.W. (1992). HLA-DR,DQ sequence polymorphisms in Polynesians, Micronesians, and Javanese. Hum Immunol *34*, 153-161.

Garcia-Ortiz, J.E., Sandoval-Ramirez, L., Rangel-Villalobos, H., Maldonado-Torres, H., Cox, S., Garcia-Sepulveda, C.A., Figuera, L.E., Marsh, S.G., Little, A.M., Madrigal, J.A.*, et al.* (2006). High-resolution molecular characterization of the HLA class I and class II in the Tarahumara Amerindian population. Tissue Antigens *68*, 135-146.

Gray, R.D., and Jordan, F.M. (2000). Language trees support the express-train sequence of Austronesian expansion. Nature *405*, 1052-1055.

Green, R.C. (2000). A range of disciplines support a dual origin for the bottle gourd in the Pacific. Journal of Polynesian Society, 191-197.

Hagelberg, E., and Clegg, J.B. (1993). Genetic polymorphisms in prehistoric Pacific islanders determined by analysis of ancient bone DNA. Proc Biol Sci *252*, 163-170.

Hagelberg, E., Kayser, M., Nagy, M., Roewer, L., Zimdahl, H., Krawczak, M., Lio, P., and Schiefenhovel, W. (1999). Molecular genetic evidence for the human settlement of the Pacific: analysis of mitochondrial DNA, Y chromosome and HLA markers. Philos Trans R Soc Lond B Biol Sci *354*, 141-152.

Hagelberg, E., Quevedo, S., Turbon, D., and Clegg, J.B. (1994). DNA from ancient Easter Islanders. Nature *369*, 25-26.

Hedrick, P., and Kumar, S. (2001). Mutation and linkage disequilibrium in human mtDNA. Eur J Hum Genet *9*, 969-972.

Hurles, M.E., Maund, E., Nicholson, J., Bosch, E., Renfrew, C., Sykes, B.C., and Jobling, M.A. (2003). Native American Y chromosomes in Polynesia: the genetic impact of the Polynesian slave trade. Am J Hum Genet *72*, 1282-1287.

Irwin, G. (1992). Prehistoric Exploration and Colonisation of the Pacific (New York, Cambridge University Press).

Jennings, J.D., ed. (1979). The Prehistory of Polynesia (Canberra, Australian National University Press).

Jobling, M.A., Hurles, M., and Tyler-Smith, C. (2004). Human evolutionary genetics : origins, peoples & disease (New York, Garland Science).

Kayser, M., Brauer, S., Weiss, G., Underhill, P.A., Roewer, L., Schiefenhovel, W., and Stoneking, M. (2000). Melanesian origin of Polynesian Y chromosomes. Curr Biol *10*, 1237-1246.

Kuby, J. (1997). Immunology, 3 edn (New York, Freeman).

Layrisse, Z., Guedez, Y., Dominguez, E., Paz, N., Montagnani, S., Matos, M., Herrera, F., Ogando, V., Balbas, O., and Rodriguez-Larralde, A. (2001). Extended HLA haplotypes in a Carib Amerindian population: the Yucpa of the Perija Range. Hum Immunol *62*, 992-1000.

Lie, B.A., Dupuy, B.M., Spurkland, A., Fernandez-Vina, M.A., Hagelberg, E., and Thorsby, E. (2007). Molecular genetic studies of natives on Easter Island: evidence of an early European and Amerindian contribution to the Polynesian gene pool. Tissue Antigens *69*, 10-18.

Mack, S.J., Bugawan, T.L., Moonsamy, P.V., Erlich, J.A., Trachtenberg, E.A., Paik, Y.K., Begovich, A.B., Saha, N., Beck, H.P., Stoneking, M.*, et al.* (2000). Evolution of Pacific/Asian populations inferred from HLA class II allele frequency distributions. Tissue Antigens *55*, 383-400.

Maitland, K., Bunce, M., Harding, R.M., Barnardo, M.C., Clegg, J.B., Welsh, K., Bowden, D.K., and Williams, T.N. (2004). HLA class-I and class-II allele frequencies and two-locus haplotypes in Melanesians of Vanuatu and New Caledonia. Tissue Antigens *64*, 678-686.

Martinez-Arends, A., Layrisse, Z., Arguello, R., Herrera, F., Montagnani, S., Matos, M., Ross, J., Dunn, P., Marsh, S.G., and Madrigal, J.A. (1998). Characterization of the HLA class I genotypes of a Venezuelan Amerindian group by molecular methods. Tissue Antigens *52*, 51-56.

Melton, T., Peterson, R., Redd, A.J., Saha, N., Sofro, A.S., Martinson, J., and Stoneking, M. (1995). Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. Am J Hum Genet *57*, 403-414.

O'Shaughnessy, D.F., Hill, A.V., Bowden, D.K., Weatherall, D.J., and Clegg, J.B. (1990). Globin genes in Micronesia: origins and affinities of Pacific Island peoples. Am J Hum Genet *46*, 144-155.

Oppenheimer, S.J., and Richards, M. (2001). Polynesian origins. Slow boat to Melanesia? Nature *410*, 166-167.

Petzl-Erler, M.L., Gorodezky, C., Layrisse, Z., Klitz, W., Fainboim, L., Vullo, C., Bodmer, J.G., Egea, E., Navarrete, C., Infante, E.*, et al.* (1997). Anthropolgy report for Region Latin-America: Amerindian and admixed populations. Paper presented at: HLA Genetic diversity of HLA; Functional and Medical Implication (EDK Medical and Scientific International Publisher).

Redd, A.J., and Stoneking, M. (1999). Peopling of Sahul: mtDNA variation in aboriginal Australian and Papua New Guinean populations. Am J Hum Genet *65*, 808-828.

Redd, A.J., Takezaki, N., Sherry, S.T., McGarvey, S.T., Sofro, A.S., and Stoneking, M. (1995). Evolutionary history of the COII/tRNA^{Lys} intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. Mol Biol Evol *12*, 604-615.

Relethford, J.H. (2003). Reflections of our past : how human history is revealed in our genes (Boulder, Colo., Westview Press).

Sayer, D.C., Goodridge, D.M., and Christiansen, F.T. (2004a). Assign 2.0: software for the analysis of Phred quality values for quality control of HLA sequencing-based typing. Tissue Antigens *64*, 556-565.

Sayer, D.C., Whidborne, R., De Santis, D., Rozemuller, E.H., Christiansen, F.T., and Tilanus, M.G. (2004b). A multicenter international evaluation of single-tube amplification protocols for sequencing-based typing of HLA-DRB1 and HLA-DRB3,4,5. Tissue Antigens *63*, 412-423.

Severson, L.D., Crews, D.E., and Lang, R.W. (1997). Application of SSP/ARMS to HLA class I loci in Samoans. Paper presented at: HLA Genetic diversity of HLA; Functional and Medical Implication (EDK Medical and Scientific International Publisher).

Su, B., Jin, L., Underhill, P., Martinson, J., Saha, N., McGarvey, S.T., Shriver, M.D., Chu, J., Oefner, P., Chakraborty, R.*, et al.* (2000). Polynesian origins: insights from the Y chromosome. Proc Natl Acad Sci U S A *97*, 8225-8228.

Terrell, J.E. (1988). History as a family tree, history as an entangled bank: constructing images and interpretations of prehistory in the South Pacific. Antiquity *62*, 642-657.

Tracey, M.C., and Carter, J.M. (2006). Class II HLA allele polymorphism: DRB1, DQB1 and DPB1 alleles and haplotypes in the New Zealand Maori population. Tissue Antigens *68*, 297-302.

Velickovic, Z.M., Delahunt, B., and Carter, J.M. (2002). HLA-DRB1 and HLA-DQB1 polymorphisms in Pacific Islands populations. Tissue Antigens *59*, 397-406.

Watkins, D.I., McAdam, S.N., Liu, X., Strang, C.R., Milford, E.L., Levine, C.G., Garber, T.L., Dogon, A.L., Lord, C.I., Ghim, S.H.*, et al.* (1992). New recombinant HLA-B alleles in a tribe of South American Amerindians indicate rapid evolution of MHC class I loci. Nature *357*, 329-333.

Wickler, S., and Spriggs, M. (1988). Pleistocene human occupation of the Solomon Islands, Melanesia. Antiquity *62*, 703-706.